A New Training Protocol for Channel State Estimation in Wireless Relay Networks

Cenk M. Yetis*, *Member, IEEE,* and Ahmet H. Kayran, *Senior Member, IEEE*

Abstract

The accuracy of channel state information (CSI) is critical for improving the capacity of wireless networks. In this paper, we introduce a training protocol for wireless relay networks that uses channel estimation and feedforwarding methods. The feedforwarding method is the distinctive feature of the proposed protocol. As we show, each relay feedforwards the imperfect CSI to the destination in a way that provides a higher network capacity and a faster transfer of the CSI than the existing protocols. In addition, we show the importance of the effective CSI accuracy on the wireless relay network capacity by comparing networks with the perfect effective CSI, imperfect effective CSI, and noisy imperfect effective CSI available at the destination.

Index Terms

Relay networks, imperfect channel state information (CSI), channel estimation, effective SNR, capacity, training.

I. INTRODUCTION

Using training signals is a widely used channel estimation method for obtaining the imperfect channel state information (CSI) in a wireless network that uses time division duplexing [1]. The information theoretic analysis of training based methods was initiated in [2], where the authors address the question of how much of the system resources should be spent for training. In [2], the system resources including

C. M. Yetis is with Informatics Institute, Satellite Remote Sensing and Communication, Istanbul Technical University, Maslak, Istanbul, 34469, TURKEY. Email: cenkmyetis@yahoo.com. The author is supported in part by The Scientific and Technological Research Council of Turkey (TUBITAK). The author is on leave at University of California, Irvine.

A. H. Kayran is with Department of Electronics and Communications Engineering, Istanbul Technical University, Maslak, Istanbul, 34469, TURKEY. Email: kayran@itu.edu.tr.

the training signal length, power, and structure are optimized by maximizing the effective signal-to-noise ratio (eSNR), which eventually maximizes the channel capacity.

Recently, there has been a significant amount of research interest in wireless relay networks, which improve the capacity and coverage over traditional point-to-point wireless networks [3]–[9]. In addition, relays are effective in dealing with the non-line-of-sight problem. Furthermore, relays do not require a backhaul architecture leading to a more cost efficient and a less complex design [10].

In this paper, we introduce a training protocol for wireless relay networks, which provides a higher eSNR than the existing training protocols. While the CSI should be accurately estimated to provide a higher eSNR as will be shown in the following sections, the CSI should also be made available at the destination in a timely manner. Our protocol also provides a faster transfer of the CSI than the existing training protocols.

The organization of this paper is as follows. In Section II, we summarize the channel model and the basic principles related to the paper. In Section III, we show the superiority of our protocol over the existing protocols for a single antenna relay network. Section IV covers the multiple-antenna results of our protocol, where the noisy imperfect effective CSI is available at the destination by the feedforwarding method. Then, we compare our protocol with two challenging schemes where the perfect effective CSI and the imperfect effective CSI are conceptually available (i.e., genie aided) at the destination. The comparison of these cases shows the importance of the effective CSI accuracy on the eSNR.

II. CHANNEL MODEL AND BASIC PRINCIPLES

This paper builds upon the concepts of [2], where the training based channel estimation is studied from an information theoretic approach. In this section, we outline the channel model, the training based channel estimation, the capacity of a network with a training phase, the worst case noise theorem, and the basics of a single antenna relay network.

A. General Features and Assumptions

We obtain a lower bound on the exact eSNR of wireless relay networks with channel estimation error by using the worst case noise theorem [2]. We assume that the number of relays is large ($K \gg 1$), which clearly shows the performance differences between the protocols compared in this work. In the wireless relay network literature, scaling number of relays $(K \to \infty)$ assumption is commonly used [6,7,11,12], which leads to the same eSNR for all the protocols (which is also the case in [6]). Further performance gain can be achieved by carefully constructing a power allocation policy. In addition, path

loss and shadowing models can be incorporated for a more realistic scenario. However, such scenarios are beyond the scope of this work. We apply the aforementioned features for all the protocols. During the comparisons, these features do not provide any bias for any of the protocols.

B. Channel Model

We use a block fading channel model where the channel is constant for a time interval T and takes an independent value for each time interval T. We assume that the transmissions of the training and the data signals are completed within the interval $T = T_\tau + T_d$, where T_x denotes the time spent for the training signal and the data signal when $x = \tau$ and $x = d$, respectively.

During the training (data transmission) phase, the received training (data) signal is as follows:

$$\mathbf{y}_x = \sqrt{\frac{\rho_x}{M}}\mathbf{s}_x\mathbf{H} + \mathbf{v}_x,$$

where $\mathbf{H} \in \mathbb{C}^{M\times N}$ is the $M \times N$ complex Gaussian channel matrix, M and N denote the number of transmit and receive antennas, respectively, $\mathbf{s}_x \in \mathbb{C}^{T_x\times M}$ denotes the transmitted training (data) signal sent in T_x time samples, $\mathbf{v}_x \in \mathbb{C}^{T_x\times N}$ is the additive complex Gaussian noise, and ρ_x is the SNR during the training (data transmission) phase.

C. Communication by using the training based channel estimation

There are two phases in a communication session that incorporates training based channel estimation. The first phase is the training phase where the channel estimation is achieved and the second phase is the data transmission phase. In the training phase, the source transmits a training signal, which is known to both source and destination. Therefore, the destination estimates the channel by using the training signal $\mathbf{s}_\tau$ and the received training signal $\mathbf{y}_\tau$. The training based channel estimate $\widehat{\mathbf{H}}$ can be represented as a function of $\mathbf{y}_\tau$ and $\mathbf{s}_\tau$:

$$\widehat{\mathbf{H}} = f(\mathbf{y}_\tau, \mathbf{s}_\tau). \tag{1}$$

In the data transmission phase, the source transmits its data signal. Therefore, the destination estimates the data signal by using the received data signal $\mathbf{y}_d$ and the channel estimate $\widehat{\mathbf{H}}$, which again can be represented as $\widehat{\mathbf{s}}_d = f(\mathbf{y}_d, \widehat{\mathbf{H}})$. Note that the data signal is estimated by throwing away some information (i.e., the channel estimate $\widehat{\mathbf{H}}$ is treated as if it were the actual channel) and the residual in the received data signal due to the channel estimation error is treated as noise:

$$\mathbf{y}_d = \sqrt{\frac{\rho_d}{M}}\mathbf{s}_d\widehat{\mathbf{H}} + \mathbf{v}'_d, \tag{2}$$

where $\mathbf{v}'_d = \sqrt{\frac{\rho_d}{M}}\mathbf{s}_d\widetilde{\mathbf{H}} + \mathbf{v}_d$ is the overall noise. Note that the equality $\mathbf{H} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}$ holds and $\sqrt{\frac{\rho_d}{M}}\mathbf{s}_d\widetilde{\mathbf{H}}$ is the residue due to the channel estimation error.

A typical way to obtain the CSI is to use the minimum mean square error (MMSE) channel estimation [2], which can be written as

$$\widehat{\mathbf{H}} = \sqrt{\frac{M}{\rho_\tau}}\left(\frac{M}{\rho_\tau}\mathbf{I}_M + \mathbf{s}_\tau^*\mathbf{s}_\tau\right)^{-1}\mathbf{s}_\tau^*\mathbf{y}_\tau,$$

where $\mathbf{I}_M$ is an $M \times M$ identity matrix and $(.)^*$ denotes the conjugate transpose. It is clearly seen that the channel estimation is a function of $\mathbf{y}_\tau$ and $\mathbf{s}_\tau$ as indicated in (1).

MMSE yields tractable results since it is an unbiased estimator: $\mathrm{E}\{\widehat{\mathbf{H}}\} = \mathrm{E}\{\mathbf{H}\}$, where $\mathrm{E}\{.\}$ denotes the expectation operator. In addition, the channel estimate and the channel estimation error are uncorrelated: $\mathrm{E}\{\widehat{\mathbf{H}}\widetilde{\mathbf{H}}^*\} = \mathbf{0}$. As an example, the variance of the channel is simply obtained as follows: $\sigma_H^2 = \sigma_{\widehat{H}}^2 + \sigma_{\widetilde{H}}^2$, where $\sigma_{\widehat{H}}^2$ and $\sigma_{\widetilde{H}}^2$ are the variances of the channel estimate and channel estimation error, respectively $\big($i.e., $\sigma_H^2 = \frac{\mathrm{Tr}\{\mathrm{E}\{\mathbf{H}\mathbf{H}^*\}\}}{MN}$, where $\mathrm{Tr}\{.\}$ denotes the trace operator$\big)$.

D. The capacity of a wireless network with a training phase

The capacity of a wireless network with a training phase is as follows [2]:

$$C_\tau = \sup_{p_{s_d}(.),\mathrm{Tr}\left\{\mathbf{R}_{s_d}\right\}\le M} I\left(\mathbf{y}_\tau, \mathbf{s}_\tau, \mathbf{y}_d; \mathbf{s}_d\right),$$

where $p_{s_d}(.)$ denotes the signal distribution, and I denotes the mutual information. Therefore, the capacity is the supremum of the mutual information between the known signals ($\mathbf{y}_\tau$, $\mathbf{s}_\tau$ and $\mathbf{y}_d$) and the unknown signal ($\mathbf{s}_d$) over the distribution $p_{s_d}(.)$ under the power constraint $\mathrm{Tr}\left\{\mathbf{R}_{s_d}\right\} \le M$, where $\mathbf{R}_{s_d}$ denotes the correlation matrix of the data signal.

In [2], the following lower bound is obtained by using the worst case noise theorem:

$$C_\tau \ge C_{\text{worst}} = \frac{T - T_\tau}{T}\mathrm{E}\left\{\log\left|\mathbf{I}_N + \rho_{\text{eff}}\frac{\overline{\mathbf{H}}^*\overline{\mathbf{H}}}{M}\right|\right\},$$

where $|.|$ denotes the magnitude operator, $\overline{\mathbf{H}} = \frac{\widehat{\mathbf{H}}}{\sigma_{\widehat{H}}}$ is the normalized channel estimate, and $\rho_{\text{eff}} = \frac{\rho_d}{\sigma_{v'_d}^2}\sigma_{\widehat{H}}^2$ is the eSNR, where $\sigma_{v'_d}^2$ is the variance of the overall noise in (2).

E. The worst case noise theorem

The worst case noise theorem was introduced for a multiple access SISO wireless network in [13]. This theorem was extended to a point-to-point MIMO wireless network in [2]. The two properties of the overall noise $\mathbf{v}'_d$ in (2) that lead to untractable results are 1) $\mathbf{v}'_d$ is possibly not Gaussian, and 2) it is

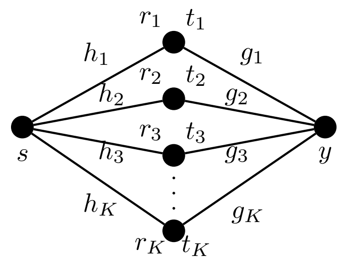

Figure 1. A single antenna wireless relay network.

dependent on the data signal. In [2], it is shown that replacing $\mathbf{v}'_d$ with a Gaussian noise that has the same power as $\mathbf{v}'_d$ has the worst effect on the channel capacity. In other words, for a Gaussian channel and a Gaussian data signal, a Gaussian noise leads to the worst eSNR (thus the worst channel capacity). First of all, this approach yields tractable results. Second, it enables us to optimize the system parameters for the worst case scenario [2]. The condition to replace $\mathbf{v}'_d$ is that the data signal and the overall noise must be uncorrelated given the CSI at the destination: $\mathrm{E}\{\mathbf{s}_d^* \mathbf{v}'_d | \widehat{\mathbf{H}}\} = 0$ (for further details, see Theorem 1 in [2] and Theorem 4.0.2 in [14]).

F. Single antenna wireless relay networks

In this subsection, we present the generic setup and the generic received signal at the destination for all protocols that will be compared in the next section. Due to brevity and clarity concerns of the paper, the comparison is carried out for a specific case where we assume $M = N = 1$ and $T_d = T_\tau = 1$. For other cases, the advantage of our protocol over the other protocols would still be observed.

In Fig. 1, a wireless relay network where each node has a single antenna is shown. The received signal at each relay is as follows:

$$r_k = \sqrt{\rho_s} s h_k + n_k, \; k = 1, 2, ..., K, \tag{3}$$

where s denotes the transmitted complex Gaussian data signal with zero mean and σ_s^2 variance, $CN(0, \sigma_s^2)$. h_k denotes the $CN(0, \sigma_{h_k}^2)$ backward channel between the source and the k^{th} relay, n_k denotes the $CN(0, \sigma_{n_k}^2)$ additive noise at the k^{th} relay, and ρ_s denotes the power spent by the source.

After each relay receives the signal (3), each relay applies the matched filtering (MF) with the backward channel estimate $\widehat{h}_k^*$:

$$u_k = r_k \widehat{h}_k^*,$$

where $\widehat{h}_k$ is the MMSE estimate of h_k. Note that

$$\mathrm{E}\left\{|u_k|^2\right\} = \left(\rho_s \left(\sigma_{\widehat{h}_k}^2 + \sigma_{\tilde{h}_k}^2\right) \sigma_s^2 + \sigma_{n_k}^2\right) \sigma_{\widehat{h}_k}^2,$$

where we have used the following equality from [15]

$$\mathrm{E}\left\{\left|\widehat{h}_k\right|^4\right\} = 2\sigma^4_{\widehat{h}_k}. \tag{4}$$

Then, each relay applies MF with the forward channel estimate $\widehat{g}^*_k$. Note that the equality $h_k = \widehat{h}_k + \widetilde{h}_k$ holds for the single antenna case as well due to the MMSE channel estimation (similarly for g_k). Considering the power constraint $\mathrm{E}\{|t_k|^2\} = 1$ at each relay, the transmitted signal from the k^{th} relay is as follows:

$$t_k = \frac{r_k \widehat{h}^*_k \widehat{g}^*_k}{\sqrt{\left(\rho_s\left(\sigma^2_{\widehat{h}_k} + \sigma^2_{h_k}\right)\sigma^2_s + \sigma^2_{n_k}\right)\sigma^2_{\widehat{h}_k}\sigma^2_{\widehat{g}_k}}}.$$

Finally, the received signal at the destination can be expressed as

$$y = y^{\text{sig}} + v, \tag{5}$$

where

$$\begin{aligned} y^{\text{sig}} &= \sum_{k=1}^{K} \sqrt{\rho_k} s h_k \widehat{h}^*_k \widehat{g}^*_k g_k, \\ v &= \sum_{k=1}^{K} \sqrt{\frac{\rho_k}{\rho_s}} n_k \widehat{h}^*_k \widehat{g}^*_k g_k + z, \text{ and} \end{aligned}$$

$$\rho_k = \frac{\rho_R \rho_s}{\left(\rho_s\left(\sigma^2_{\widehat{h}_k} + \sigma^2_{h_k}\right)\sigma^2_s + \sigma^2_{n_k}\right)\sigma^2_{\widehat{h}_k}\sigma^2_{\widehat{g}_k}}.$$

Here, ρ_R and z are the power spent by each relay and the noise at the destination, respectively, and y^{sig} and v are the received desired signal and the overall noise at the destination, respectively. Note that the effective CSI in (5) is as follows:

$$\sum_{k=1}^{K} h_k \widehat{h}^*_k \widehat{g}^*_k g_k. \tag{6}$$

If the destination knows the effective CSI (6), then the eSNR is obtained as follows:

$$\rho_{\text{eff}} = \frac{\mathrm{E}\left\{\left|y^{\text{sig}}\right|^2\right\}}{\mathrm{E}\left\{|v|^2\right\}}.$$

III. TRAINING PROTOCOLS IN SINGLE ANTENNA WIRELESS RELAY NETWORKS

In this section, we compare the classical protocol (P1) [2], the new protocol (P2) [16], and our protocol (P3). The key point is that the protocol that estimates the effective CSI (6) more accurately would provide a higher eSNR (thus a higher capacity). The protocols are briefly explained in the sequel.

P1: The source and the relays send the training signals to the relays and the destination, respectively. Then, the relays estimate the backward channels and feedforward the backward channel estimates to the destination. Therefore, the destination knows the imperfect CSI $\widehat{h}_k$ and $\widehat{g}_k$ $(k = 1, 2, \cdots, K)$.

P2: Only the source sends the training signals to the relays. Then, the relays simply feedforward the received training signals to the destination. Finally, the destination estimates the product (compound) of the backward and forward channels. Therefore, the destination knows the imperfect CSI $\widehat{a}_k$ $(k = 1, 2, \cdots, K)$, where $a_k = h_k g_k$.

P3: The source and the destination send the training signals to the relays. Each relay estimates the backward and forward channels, and then each relay feedforwards the CSI $A_k = \widehat{h}_k \widehat{h}_k^* \widehat{g}_k^*$ to the destination. Therefore, the destination has the noisy imperfect effective CSI $\sum_{k=1}^{K} A_k g_k + z^f$, where z^f is the noise during the feedforwarding phase.

The training protocols are also summarized in Table I. Note that we omit the subindex τ to denote the training signals for notational simplicity and denote the training signals sent from the source, relay, and destination by s_s, s_{R_k}, and s_D, respectively. In addition, for simplicity, only one relay is shown in the table. Finally, y_k denotes the received training signal at the k^{th} relay. Note that the horizontal axis in the table can be considered as a time line. For example, consider the protocol P1. The source sends the training signal to the k^{th} relay in the first time slot. Then, the k^{th} relay sends the training signal to the destination in the second time slot. Later, the k^{th} relay sends the backward channel estimate, which is obtained in the first time slot, to the destination in the third time slot.

A. Classical protocol (P1)

Since the destination knows each $\widehat{h}_k$ and $\widehat{g}_k$ $(k = 1...K)$, we can rewrite (5) as follows:

$$\begin{aligned} y &= \sqrt{\rho}\sum_{k=1}^{K} s\widehat{h}_k \widehat{h}_k^* \widehat{g}_k^* \widehat{g}_k \\ &\quad + \sqrt{\rho}\sum_{k=1}^{K} s\left(\widehat{h}_k \widehat{h}_k^* \widehat{g}_k^* \widetilde{g}_k + \widetilde{h}_k \widehat{h}_k^* \widehat{g}_k^* g_k\right). \end{aligned} \tag{7}$$

Table I

SUMMARY FOR THE TRAINING PROTOCOLS P1, P2, AND P3.

	Training protocol	Destination CSI
P1	S $\xrightarrow{s_s}$ $\mathrm{R_k}$ $\xrightarrow{s_{R_k}}$ $\xrightarrow{\widehat{h}_k}$ D	$\widehat{h}_k, \widehat{g}_k, k = 1, 2, ..., K$
P2	S $\xrightarrow{s_s}$ $\mathrm{R_k}$ $\xrightarrow{y_k}$ D	$\widehat{a}_k, k = 1, 2, ..., K$, where $a_k = h_k g_k$
P3	S $\xrightarrow{s_s}$ $\mathrm{R_k}$ $\xleftarrow{s_D}$ $\xrightarrow{A_k}$ D	$\sum_{k=1}^{K} A_k g_k + z^f$

Therefore, we obtain the following eSNR:

$$\rho_{\text{eff,P1}} = \frac{(K+3)\,\sigma_{\widehat{h}}^2 \sigma_{\widehat{g}}^2}{2\sigma_{\widehat{h}}^2 \sigma_{\widetilde{g}}^2 + \sigma_{\widetilde{h}}^2 \left(\sigma_g^2 + \sigma_{\widehat{g}}^2\right)}. \tag{8}$$

Note that the same overall noise exists for all protocols. Since we aim to compare the protocols, we neglect the overall noise in (5). We also omit the simple proof that the data signal and the overall noise are uncorrelated given CSI at the destination for the sake of brevity. We only prove this condition in subsection IV-A, where we apply our protocol to a multi-antenna wireless relay network.

B. New protocol (P2)

Since the destination knows each $\widehat{a}_k$ $(k = 1...K)$, we can rewrite (5) as

$$y = \sqrt{\rho}\sum_{k=1}^{K} s\widehat{a}_k \widehat{a}_k^* + \sqrt{\rho}\sum_{k=1}^{K} s\widetilde{a}_k \widehat{a}_k^*, \tag{9}$$

where $\widehat{a}_k = \widehat{h}_k \widehat{g}_k$.

Then, the eSNR is obtained as follows:

$$\rho_{\text{eff,P2}} = \frac{\mathrm{E}\left\{\sum\limits_{k,m=1}^{K} |\widehat{a}_k|^2 \,|\widehat{a}_m|^2\right\}}{\mathrm{E}\left\{\sum\limits_{k=1}^{K} |\widetilde{a}_k \widehat{a}_k|^2\right\}} = \frac{(K+1)\sigma_{\widehat{a}}^2}{\sigma_{\widetilde{a}}^2}. \tag{10}$$

As seen from (7) and (9), the assumption $\widehat{a}_k = \widehat{h}_k \widehat{g}_k$ leads to the same received signal for P1 and P2. However, assuming that the destination knows each compound channel estimate $(\widehat{a}_k)$ or each channel estimate individually $(\widehat{h}_k$ and $\widehat{g}_k)$ leads to different statistical values and therefore, leads to different eSNRs. In order to obtain (8), we need to use the property (4) since we encounter the fourth-order

moments. On the other hand, we do not encounter any such moments in order to obtain (10) since we assume that the destination knows the compound channel estimates.

We use the results obtained in [17] for determining the variances of the compound channel estimate ($\sigma^2_{\widehat{a}_k}$) and error ($\sigma^2_{\widetilde{a}_k}$) in (10). In [17], it is shown that the variance of the product of two real Gaussian random variables, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, approaches to $\sigma_1^2\sigma_2^2$ if $\frac{\mu_1}{\sigma_1}$ and $\frac{\mu_2}{\sigma_2}$ approach to zero. Note that this result also holds for two complex Gaussian random variables. As a result, we easily obtain the variances (i.e., $\sigma^2_{a_k} = \sigma^2_{h_k}\sigma^2_{g_k}$).

First of all, the assumption $\widehat{a}_k = \widehat{h}_k\widehat{g}_k$ enables us to obtain a comparable result to the other protocols' results and *directly* show the advantage of knowing the product of channel estimates ($\widehat{h}_k\widehat{g}_k$) over knowing the individual channel estimates ($\widehat{h}_k, \widehat{g}_k$). Second, our assumption enables us to obtain a tractable result since the orthogonality of $\widehat{a}_k$ and $\widetilde{a}_k$ is still preserved: the equalities $\sigma^2_{a_k} = \sigma^2_{\widehat{a}_k} + \sigma^2_{\widetilde{a}_k}$ and $a_k = \widehat{a}_k + \widetilde{a}_k$ still hold.

Therefore, we obtain the following eSNR:

$$\rho_{\text{eff,P2}} = \frac{(K+1)\sigma^2_{\widehat{h}}\sigma^2_{\widehat{g}}}{\sigma^2_{\widehat{h}}\sigma^2_{\widetilde{g}} + \sigma^2_{\widetilde{h}}\sigma^2_{g}}.$$

C. Our protocol (P3)

The following signal is received at the destination after each relay feedforwards CSI $A_k = \widehat{h}_k\widehat{h}_k^*\widehat{g}_k^*$ to the destination:

$$\sum_{k=1}^{K}\sqrt{\rho'_{R,f}}\widehat{h}_k\widehat{h}_k^*\widehat{g}_k^*g_k + z^f,$$

where $\rho'_{R,f}$ is the normalized power spent by each relay, $\rho'_{R,f} = \frac{\rho_{R,f}}{\text{var}\{A_k g_k\}}$. var$\{.\}$ denotes the power of a vector or a scalar. z^f is the noise during the feedforwarding phase. Assuming that the destination knows $\rho'_{R,f}$ noiselessly, the destination has the following effective CSI:

$$\sqrt{\rho}\left(\sum_{k=1}^{K}\widehat{h}_k\widehat{h}_k^*\widehat{g}_k^*g_k + \frac{z^f}{\sqrt{\rho'_{R,f}}}\right). \tag{11}$$

Since the destination knows the above effective CSI, we can rewrite the received signal (5) as follows:

$$y = y^{\text{sig}} + v,$$

where

$$y^{\text{sig}} = \sqrt{\rho}\left(\sum_{k=1}^{K}\widehat{h}_k\widehat{h}_k^*\widehat{g}_k^*g_k + \frac{z^f}{\sqrt{\rho'_{R,f}}}\right)s \text{ and}$$

$$v = \sqrt{\rho}\left(\sum_{k=1}^{K}\widetilde{h}_k\widehat{h}_k^*\widehat{g}_k^*g_k - \frac{z^f}{\sqrt{\rho'_{R,f}}}\right)s.$$

y^{sig} and v are the received noisy desired signal and the overall noise at the destination, respectively. Note that the intersymbol interference is critical for our protocol. Due to (11), our protocol relies on perfect synchronization at the destination, whereas the other protocols do not.

By neglecting the feedforwarding noise due to the large K assumption, we obtain the following eSNR:

$$\rho_{\text{eff,P3}} = \frac{2\sigma_{\widehat{h}}^2\left(\sigma_{\widehat{g}}^2+\sigma_{g}^2\right)+(K-1)\,\sigma_{\widehat{h}}^2\sigma_{\widehat{g}}^2}{\sigma_{\widetilde{h}}^2\left(\sigma_{\widehat{g}}^2+\sigma_{g}^2\right)}.$$

It is not immediately obvious from the previous results that the inequality $\rho_{\text{eff,P3}} > \rho_{\text{eff,P2}} > \rho_{\text{eff,P1}}$ holds. However, we easily reach to this conclusion by numerically averaging the eSNRs over the backward and forward channel estimation variances for a fixed value of K. For example, consider the case when $K = 7$. If we average the eSNRs by changing the estimated backward and forward channel variances discretely from 0.1 to 0.9 in 0.1 increments, we obtain 8.3, 3.92, and 2.67 for P3, P2, and P1, respectively.

D. The worst case noise comparison

By using the worst case noise theorem, the following capacity lower bounds can be obtained:

$$C_{\text{P}i} \geq C_{\text{P}i\text{-worst}} = \frac{1}{2}\text{E}\{I_{\text{P}i}\},\ i = 1, 2, 3, \tag{12}$$

where

$$I_{\text{P1}} = \log_2\left(1+\frac{A^2}{B\left(C+2\sigma_{\widehat{h}}^2\sigma_{\widetilde{g}}^2\right)}\right),$$

$$I_{\text{P2}} = \log_2\left(1+\frac{A^2}{BC'}\right), \text{ and}$$

$$I_{\text{P3}} = \log_2\left(1+\frac{A^2+\left|\sum_{k=1}^{K}\left|\widehat{h}_k\right|^2\widehat{g}_k^*\widetilde{g}_k\right|^2}{BC}\right).$$

In (12),

$$A = \sum_{k=1}^{K}\left|\widehat{h}_k\right|^2|\widehat{g}_k|^2, B = K\sigma_{\widehat{h}}^2\sigma_{\widehat{g}}^2,$$

$$C = \left(\sigma^2_{\tilde{h}} + \sigma^2_n\right)\left(\sigma^2_{\hat{g}} + \sigma^2_g\right), \text{and}$$

$$C' = \sigma^2_{\hat{h}}\sigma^2_{\tilde{g}} + \sigma^2_{\tilde{h}}\sigma^2_g + \left(\sigma^2_{\hat{g}} + \sigma^2_g\right)\sigma^2_n.$$

For the sake of brevity, we omit the simple proofs showing that the data signal and the overall noise are uncorrelated given CSI at the destination. Note that the numerators in (12) are the known CSI at the destination. Therefore, there are no expectation operators in the numerators.

It is not obvious from the capacity lower bounds in (12) that the inequality $C_{\text{P3-worst}} > C_{\text{P2-worst}} \geq C_{\text{P1-worst}}$ holds. However, this result can be clearly deduced from Table II, where the capacity results for P1, P2, and P3 are shown for $\sigma^2_{\hat{h}}, \sigma^2_{\hat{g}} \in \{0.1, 0.5, 0.9\}$. The right most values in Table II are for no channel estimation error cases (i.e., $\sigma^2_{\hat{h}} = \sigma^2_h = 1$ and $\sigma^2_{\hat{g}} = \sigma^2_g = 1$), which are again obtained from (12).

The simulation results in Table II are obtained by assuming that the channels have unit variances ($\sigma^2_h = \sigma^2_g = 1$). Assuming that the channels have unit variances is different than assuming that the channels have equal variances ($\sigma^2_h = \sigma^2_g$) because the former assumption is a *normalization*. e.g., $\sigma^2_h = \sigma^2_{\hat{h}} + \sigma^2_{\tilde{h}} = 1$. Therefore, a channel estimation variance that has a closer value to 1 indicates a more accurate channel estimation.

Note that the higher capacity performance of our protocol P3 is due to our feedforwarding method where the channel estimates are sent to the destination in a way that the relay network carries the perfect forward CSI (i.e., g_k in (11)). In fact, the underlying key observation is that estimating the effective CSI more accurately provides a higher capacity.

Finally, the results in Table II show that estimating the effective CSI more directly provides a higher capacity. For example, P2 provides a higher capacity than P1 because P2 estimates the effective CSI more directly than P1. While P1 estimates the backward and the forward channels individually, P2 estimates the product of the backward and forward channels. For the no channel estimation error case, our protocol does not have an advantage over the other protocols due to the relay network structure that carries the perfect forward CSI. However, our protocol still provides a higher capacity than the others because our protocol provides the effective CSI at the destination more directly than the other protocols. Note that the performance difference between P1 and P2 vanishes towards the opposite ends of the channel estimation error values as seen in Table II.

Finally, having more accurate backward channel estimates is more valuable than having more accurate forward channel estimates since the former leads to a higher capacity as seen in Table II. This result is due to the asymmetric nature of MF as explained in detail in the subsection IV-C.

Table II

CAPACITY RESULTS (BITS/SEC/CHANNEL USE) FOR THE TRAINING PROTOCOLS P1, P2, AND P3 ($M = N = 1$ AND $K = 7$).

(a) $C_{\text{P1-worst}}$

$\sigma_{\hat{g}}^2/\sigma_{\hat{h}}^2$	0.1	0.5	0.9	
0.9	0.14	0.52	0.84	0.99
0.5	0.1	0.37	0.56	
0.1	0.03	0.11	0.17	

(b) $C_{\text{P2-worst}}$

$\sigma_{\hat{g}}^2/\sigma_{\hat{h}}^2$	0.1	0.5	0.9	
0.9	0.17	0.58	0.87	0.99
0.5	0.12	0.42	0.64	
0.1	0.03	0.14	0.22	

(c) $C_{\text{P3-worst}}$

$\sigma_{\hat{g}}^2/\sigma_{\hat{h}}^2$	0.1	0.5	0.9	
0.9	0.26	0.82	1.26	1.4
0.5	0.21	0.71	1.11	
0.1	0.11	0.43	0.74	

E. *Fast transfer of CSI*

In this subsection, we show that our protocol also provides a faster transfer of the CSI than the existing training protocols. Consider a single antenna relay network with a time division duplexed training and assume that each training signal length is one symbol duration.

In P1, the source sends the training signal to the relays, which requires 1 symbol duration. Then, the relays send the training signals to the destination, which requires K symbol durations. As a result, $(K+1)$ symbol durations are required for the training phase. In the feedforwarding phase, each relay feedforwards the backward channel estimate to the destination. Therefore, K symbol durations are required for the feedforwarding phase. In total, $(2K + 1)$ symbol durations are required for P1 to transfer the CSI to the destination, $T_{\text{P1}} = 2K + 1$.

In P2, the source sends the training signal to the relays, which requires 1 symbol duration. Then, each relay feedforwards this training signal to the destination, which requires K symbol durations. In total, $(K + 1)$ symbol durations are required for P2, $T_{\text{P2}} = K + 1$.

In P3, the source and the destination send the training signals to the relays, which requires 2 symbol durations. Then, each relay feedforwards the CSI A_k to the destination simultaneously. Therefore, only 1 symbol duration is required for the feedforwarding phase. In total, 3 symbol durations are required for P3, $T_{\text{P3}} = 3$.

Therefore, $T_{\text{P1}} > T_{\text{P2}} \geq T_{\text{P3}}$ for $K \geq 2$. Note the superiority of our protocol over the other protocols. Moreover, our protocol is independent of the number of relays K, which is important for large K values.

IV. OUR PROTOCOL IN MULTI-ANTENNA WIRELESS RELAY NETWORKS

In this section, we apply the training protocol that we have proposed in subsection III-C to a multi-antenna wireless relay network (Scheme S2). Then, we compare S2 with its challenging schemes (S1a and S1b) in terms of eSNR.

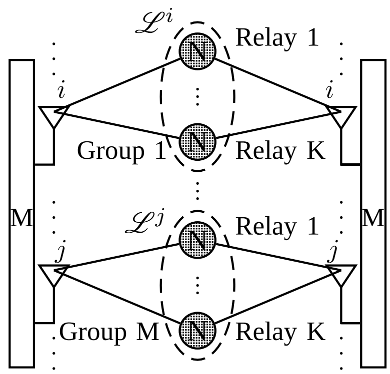

Figure 2. Relay partitioning in a multi-antenna relay network.

We consider a scenario with one source, one destination and MK relays. The source and the destination have M antennas each, while each relay has N antennas. We use the notation $M_{[1]} \times N_{[MK]} \times M_{[1]}$ to represent this network setup. We use the single antenna decoding method [6], where the designated antenna pairs of the source and the destination communicate (each source antenna communicates to a separate destination antenna) through the assigned relays as shown in Fig. 2. Assigning the relays (subgrouping the relays) is called relay partitioning [6]. In the single antenna decoding method, we assume that the x^{th} $(x = 1, 2, ..., M)$ source and the x^{th} destination antenna are assigned to each other for notational convenience. Each antenna pair is served by a subgroup of K relays, thus there are M subgroups. The subgroup that serves for the x^{th} antenna pair is denoted by $\mathscr{L}^x$ $(x = 1, 2, ..., M)$.

Considering the communication between the i^{th} antenna pair, the received signal at the k^{th} relay in the i^{th} subgroup is as follows:

$$\mathbf{r}_{ik} = \sqrt{\rho_s} \left(\mathbf{s}^i \mathbf{h}^i_{ik} + \sum_{\substack{j=1 \\ j \neq i}}^{M} \mathbf{s}^j \mathbf{h}^j_{ik} \right) + \mathbf{n}_{ik},$$
$$i = 1, 2, ..., M \text{ and } k = 1, 2, ..., K, \tag{13}$$

where ρ_s is the power spent by the source for the data transmission phase, and $\mathbf{s}^y$ is the $T_d \times 1$ data signal vector transmitted from the y^{th} antenna with independent and identically distributed (i.i.d.) $CN(0, \sigma^2_{s^y})$ entries. $\mathbf{h}^y_{xk}$ is a $1 \times N$ backward channel between the y^{th} source antenna and the k^{th} relay in the x^{th} subgroup, and has i.i.d. $CN(0, \sigma^2_{h^y_x})$ entries. $\mathbf{n}_{xk}$ is the noise at the k^{th} relay of the x^{th} subgroup and has i.i.d. $CN(0, \sigma^2_{n_x})$ entries. Note that the second term inside the parenthesis in (13) is the interference.

After each relay applies MF, the received signal at the i^{th} destination antenna is as follows:

$$\mathbf{y}^i = \sqrt{\frac{\rho_R}{N}}\left(\sum_{k,R_k\in\mathscr{L}^i}\mathbf{t}_{ik}\mathbf{g}_{ik}^i+\sum_{\substack{j=1\\ j\neq i}}^{M}\left(\sum_{k,R_k\in\mathscr{L}^j}\mathbf{t}_{jk}\mathbf{g}_{jk}^i\right)\right) +\mathbf{z}^i, i=1,2,...,M,$$

where ρ_R is the power spent by each relay, and $\mathbf{t}_{xk}$ is the $T_d\times N$ transmitted signal at the k^{th} relay in the x^{th} subgroup. The k^{th} relay is denoted by R_k $(k=1,2,...,K)$. $\mathbf{g}_{xk}^y$ is a $N\times 1$ forward channel between the y^{th} destination antenna and the k^{th} relay in the x^{th} subgroup, and has i.i.d. $CN(0,\sigma^2_{g_x^y})$ entries. $\mathbf{z}^y$ is the noise at the y^{th} destination antenna and has i.i.d. $CN(0,\sigma^2_{z^y})$ entries. The desired data signal is represented with i and the interfering signals are represented with j $(j=1,2,...,M$ and $j\neq i)$ as also seen in Fig. 3.

A. Effective SNR of scheme S2

The following lemma is useful for the derivation of the eSNR of S2. The proof of the lemma is straightforward by using the moment theorems for complex Gaussian processes [15] but tedious.

Lemma 1: Let $\mathbf{s},\mathbf{h},\mathbf{g}$, and $\mathbf{n}$ be the data, backward channel, forward channel, and noise vector, respectively. The data signal is a $T_d\times 1$ vector and has i.i.d. $CN(0,\sigma_s^2)$ entries. The backward and the forward channels are $1\times N$ and $N\times 1$ vectors, and have i.i.d. $CN(0,\sigma_h^2)$ and $CN(0,\sigma_g^2)$ entries, respectively. The noise is a $T_d\times N$ vector and has i.i.d. $CN(0,\sigma_n^2)$ entries. The bar ($\bar{\mathbf{x}}$) denotes either the vector itself ($\mathbf{x}$), estimation of the vector ($\widehat{\mathbf{x}}$), or an independent same type vector ($\mathbf{x}'$). Then, we have

- $\mathrm{var}\{\mathbf{s}\mathbf{h}\bar{\mathbf{h}}^*\mathbf{g}^*\bar{\mathbf{g}}\}=\mathrm{var}\{\mathbf{s}\}\mathrm{var}\{\mathbf{h}\bar{\mathbf{h}}^*\}\mathrm{var}\{\mathbf{g}^*\bar{\mathbf{g}}\}$,
- $\mathrm{var}\{\mathbf{n}\mathbf{h}^*\mathbf{g}^*\bar{\mathbf{g}}\}=\mathrm{var}\{\mathbf{n}\mathbf{h}^*\}\mathrm{var}\{\mathbf{g}^*\bar{\mathbf{g}}\}$,
- $\mathrm{var}\{\mathbf{h}\mathbf{h}^*\}=N(N+1)\sigma_h^4$,
- $\mathrm{var}\{\mathbf{h}\widehat{\mathbf{h}}^*\}=N\sigma^2_{\widehat{h}}(N\sigma^2_{\widehat{h}}+\sigma_h^2)$, and
- $\mathrm{var}\{\mathbf{h}\mathbf{h}'^*\}=N\sigma_h^2\sigma_{h'}^2$.

Similar results can be obtained for the forward channel.

Theorem 1: For the scheme with our proposed training protocol (S2), the eSNR is given by

$$\rho_{\text{eff,S2}}=\frac{N1}{D1+D2+D3+D4}, \tag{14}$$

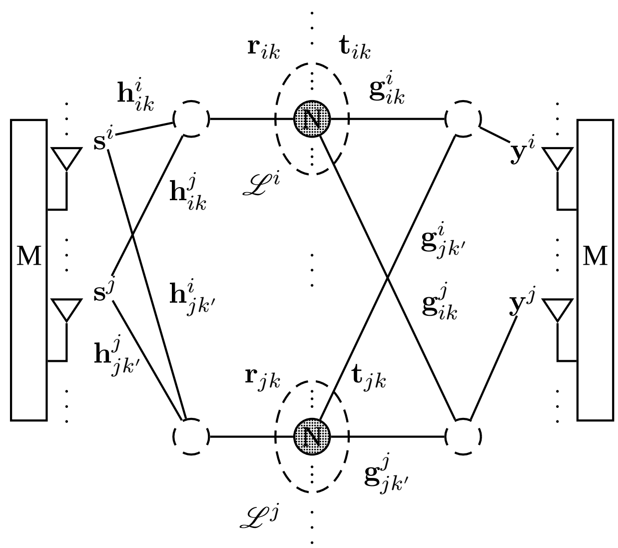

Figure 3. Notations used in a $M_{[1]} \times N_{[MK]} \times M_{[1]}$ network with relay partitioning.

where

$$N1 = \frac{\rho_s\left((N+1)\sigma^2_{\widehat{h}^i_i}\left(N\sigma^2_{\widehat{g}^i_i}+\sigma^2_{g^i_i}\right)+(K-1)N^2\sigma^2_{\widehat{h}^i_i}\sigma^2_{\widehat{g}^i_i}\right)\sigma^2_{s^i}}{A},$$

$$D1 = \frac{\rho_s\begin{pmatrix} \sigma^2_{\widetilde{h}^i_i}\left(N\sigma^2_{\widehat{g}^i_i}+\sigma^2_{g^i_i}\right)\sigma^2_{s^i} \\ +(M-1)\sigma^2_{h^j_i}\left(N\sigma^2_{\widehat{g}^i_i}+\sigma^2_{g^i_i}\right)\sigma^2_{s^j} \end{pmatrix}}{A},$$

$$D2 = \frac{\rho_s(M-1)\begin{bmatrix} \left(N\sigma^2_{\widehat{h}^j_j}+\sigma^2_{h^j_j}\right)\sigma^2_{s^j} \\ +\,(M-2)\,\sigma^2_{h^m_j}\sigma^2_{s^m}+\sigma^2_{h^i_j}\sigma^2_{s^i} \end{bmatrix}\sigma^2_{g^i_j}}{B},$$

$$D3 = \frac{\left(N\sigma^2_{\widehat{g}^i_i}+\sigma^2_{g^i_i}\right)\sigma^2_{s^j}}{A}, \quad D4 = \frac{(M-1)\sigma^2_{g^i_j}\sigma^2_{s^j}}{B}\text{ , and}$$

$$\begin{aligned} A &= \rho_s\left(\left(N\sigma^2_{\widehat{h}^i_i}+\sigma^2_{h^i_i}\right)\sigma^2_{s^i}+(M-1)\left(\sigma^2_{h^j_i}\sigma^2_{s^j}\right)\right) \\ &\quad +\sigma^2_{n_i}, \\ B &= \rho_s\Bigg(\sigma^2_{h^i_j}\sigma^2_{s^i}+\left(N\sigma^2_{\widehat{h}^j_j}+\sigma^2_{h^j_j}\right)\sigma^2_{s^j} \\ &\quad +(M-2)\sigma^2_{h^m_j}\sigma^2_{s^m}\Bigg)+\sigma^2_{n_j}. \end{aligned}$$

Proof: In S2, each relay applies MF with the imperfect backward CSI $\widehat{\mathbf{h}}^{i*}_{ik}$ after receiving the signal (13):

$$\mathbf{u}_{ik}=\mathbf{r}_{ik}\widehat{\mathbf{h}}^{i*}_{ik}. \tag{15}$$

The signal (15) is normalized to unit average power

$$\begin{aligned} \text{var}\{\mathbf{u}_{ik}\} &= \Bigg[\rho_s\Bigg(\left(N\sigma^2_{\widehat{h}^i_i}+\sigma^2_{h^i_i}\right)\sigma^2_{s^i} \\ &\quad +(M-1)\,\sigma^2_{h^j_i}\sigma^2_{s^j}\Bigg)+\sigma^2_{n_i}\Bigg]NT_d\sigma^2_{\widehat{h}^i_i} \end{aligned}$$

by using the following equalities from Lemma 1:

$$\begin{aligned} \text{var}\{\mathbf{h}^i_{ik}\widehat{\mathbf{h}}^{i*}_{ik}\} &= N\sigma^2_{\widehat{h}^i_i}\left(N\sigma^2_{\widehat{h}^i_i}+\sigma^2_{h^i_i}\right) \text{ and} \\ \text{var}\{\mathbf{h}^j_{ik}\widehat{\mathbf{h}}^{i*}_{ik}\} &= N\sigma^2_{\widehat{h}^i_i}\sigma^2_{h^j_i}. \end{aligned}$$

Then, each relay applies MF with the imperfect forward CSI $\widehat{\mathbf{g}}^{i*}_{ik}$. Considering the power constraint $\text{Tr}\left\{\text{E}\{\mathbf{t}^*_{ik}\mathbf{t}_{ik}\}\right\}\leq NT_d$ at the relays, the following signal is transmitted from the k^{th} relay in the i^{th} subgroup:

$$\mathbf{t}_{ik}=\frac{\mathbf{u}_{ik}\widehat{\mathbf{g}}^{i*}_{ik}}{\sqrt{\left[\rho_s\left(\begin{array}{c}\left(N\sigma^2_{\widehat{h}^i_i}+\sigma^2_{h^i_i}\right)\sigma^2_{s_i}\\ +(M-1)\,\sigma^2_{h^j_i}\sigma^2_{s^j}\end{array}\right)+\sigma^2_{n_i}\right]N\sigma^2_{\widehat{h}^i_i}\sigma^2_{\widehat{g}^i_i}}}.$$

Similar steps are followed for the received signal $\mathbf{r}_{jk}$ $(j=1,...,M$ and $j\neq i)$ at the k^{th} relay in the j^{th} subgroup. For this case, MF with the imperfect backward CSI $\widehat{\mathbf{h}}^j_{jk}$ and the imperfect forward CSI $\widehat{\mathbf{g}}^j_{jk}$ are applied.

Finally, the received signal at the destination is given by

$$\mathbf{y}^i=\mathbf{y}^{i,\text{sig}}+\mathbf{y}^{i,\text{int}}+\mathbf{v}^i,\ i=1,2,...,M, \tag{16}$$

where

$$
\begin{aligned}
\mathbf{y}^{i,\text{sig}} &= \mathbf{s}^i h^{i,\text{sig}}, \\
\mathbf{y}^{i,\text{int}} &= \sum_{\substack{j=1\\ j\neq i}}^{M} \left(\mathbf{s}^j h_j^{i,\text{int}} + \sum_{\substack{m=1\\ m\neq i\\ m\neq j}}^{N} \mathbf{s}^m h_{j,m}^{i,\text{int}} \right), \text{ and}
\end{aligned}
$$

$$
\begin{aligned}
h^{i,\text{sig}} &= \sqrt{\frac{\rho_i}{N}} \sum_{k, R_k \in \mathscr{L}^i} \mathbf{h}_{ik}^{i} \widehat{\mathbf{h}}_{ik}^{i*} \widehat{\mathbf{g}}_{ik}^{i*} \mathbf{g}_{ik}^{i} \\
&\quad + \sum_{\substack{j=1\\ j\neq i}}^{M} \sum_{k, R_k \in \mathscr{L}^j} \sqrt{\frac{\rho_j}{N}} \mathbf{h}_{jk}^{i} \widehat{\mathbf{h}}_{jk}^{j*} \widehat{\mathbf{g}}_{jk}^{j*} \mathbf{g}_{jk}^{i}, \qquad (16a) \\
h_j^{i,\text{int}} &= \sqrt{\frac{\rho_i}{N}} \sum_{\substack{j=1\\ j\neq i}}^{M} \sum_{k, R_k \in \mathscr{L}^j} \mathbf{h}_{ik}^{j} \widehat{\mathbf{h}}_{ik}^{i*} \widehat{\mathbf{g}}_{ik}^{i*} \mathbf{g}_{ik}^{i} \\
&\quad + \sum_{\substack{j=1\\ j\neq i}}^{M} \sum_{k, R_k \in \mathscr{L}^j} \sqrt{\frac{\rho_j}{N}} \mathbf{h}_{jk}^{i} \widehat{\mathbf{h}}_{jk}^{j*} \widehat{\mathbf{g}}_{jk}^{j*} \mathbf{g}_{jk}^{i}, \qquad (16b) \\
h_{j,m}^{i,\text{int}} &= \sum_{\substack{j=1\\ j\neq i}}^{M} \sum_{k, R_k \in \mathscr{L}^j} \sum_{\substack{m=1\\ m\neq i\\ m\neq j}}^{N} \sqrt{\frac{\rho_j}{N}} \mathbf{h}_{jk}^{m} \widehat{\mathbf{h}}_{jk}^{j*} \widehat{\mathbf{g}}_{jk}^{j*} \mathbf{g}_{jk}^{i}, \qquad (16c)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{v}^i &= \sqrt{\frac{\rho_i}{\rho_s N}} \sum_{k, R_k \in \mathscr{L}^i} \mathbf{n}_{ik} \widehat{\mathbf{h}}_{ik}^{i*} \widehat{\mathbf{g}}_{ik}^{i*} \mathbf{g}_{ik}^{i} \\
&\quad + \sum_{\substack{j=1\\ j\neq i}}^{M} \sum_{k, R_k \in \mathscr{L}^j} \sqrt{\frac{\rho_j}{\rho_s N}} \mathbf{n}_{jk} \widehat{\mathbf{h}}_{jk}^{j*} \widehat{\mathbf{g}}_{jk}^{j*} \mathbf{g}_{jk}^{i} + \mathbf{z}^i, \text{ and}
\end{aligned}
$$

$$
\rho_i = \frac{\rho_R}{\left[\rho_s \begin{pmatrix} \left(N\sigma^2_{\widehat{h}_i^i} + \sigma^2_{h_i^i} \right) \sigma^2_{s^i} \\ + (M-1)\, \sigma^2_{h_i^j} \sigma^2_{s^j} \end{pmatrix} + \sigma^2_{n_i} \right] N\sigma^2_{\widehat{h}_i^i} \sigma^2_{\widehat{g}_i^i}},
$$

$$
\rho_j = \frac{\rho_R}{\left[\rho_s \begin{pmatrix} \sigma^2_{h_j^i} \sigma^2_{s^i} \\ + \left(N\sigma^2_{\widehat{h}_j^j} + \sigma^2_{h_j^j} \right) \sigma^2_{s^j} \\ + (M-2)\, \sigma^2_{h_j^m} \sigma^2_{s^m} \end{pmatrix} + \sigma^2_{n_j} \right] N\sigma^2_{\widehat{h}_j^j} \sigma^2_{\widehat{g}_j^j}}.
$$

During the feedforwarding phase, each relay in the subgroup $\mathscr{L}^i$ $(i = 1, 2, ..., M)$ transmits $\mathbf{A}_{ik} = \widehat{\mathbf{h}}^i_{ik}\widehat{\mathbf{h}}^{i*}_{ik}\widehat{\mathbf{g}}^{i*}_{ik}$ to the destination. Therefore, the received signal at the i^{th} destination ($i = 1, 2, ..., M$) is given by

$$\sum_{k, R_k \in \mathscr{L}^i} \sqrt{\frac{\rho'_{R,f}}{N}} \widehat{\mathbf{h}}^i_{ik}\widehat{\mathbf{h}}^{i*}_{ik}\widehat{\mathbf{g}}^{i*}_{ik}\mathbf{g}^i_{ik} + \mathbf{z}^{i,f},$$

where $\rho'_{R,f} = \frac{\rho_{R,f}}{\text{var}\{\mathbf{A}_k \mathbf{g}^i_{ik}\}}$ is the normalized power spent by each relay. Assuming that destination knows $\rho'_{R,f}$ noiselessly, destination has the following effective CSI:

$$\sqrt{\frac{\rho}{N}} \sum_{k, R_k \in \mathscr{L}^i} \widehat{\mathbf{h}}^i_{ik}\widehat{\mathbf{h}}^{i*}_{ik}\widehat{\mathbf{g}}^{i*}_{ik}\mathbf{g}^i_{ik} + \sqrt{\frac{\rho}{\rho'_{R,f}}} \mathbf{z}^{i,f}. \tag{17}$$

We rearrange the received signal (16) according to the noisy imperfect effective CSI (17) as follows:

$$\mathbf{y}^i = \mathbf{y}^{i,\text{sig}'} + \overbrace{\mathbf{y}^{i,\text{int}'} + \underbrace{\mathbf{y}^{i,\text{int}} + \mathbf{v}^i}_{\mathbf{Z}^i}}^{\mathbf{V}^i}, i = 1, 2, ..., M, \tag{18}$$

where

$$\begin{aligned}
\mathbf{y}^{i,\text{sig}'} &= \mathbf{s}^i h^{i,\text{sig}'}, \ \mathbf{y}^{i,\text{int}'} = \mathbf{s}^i h^{i,\text{int}'}, \text{ and} \\
h^{i,\text{sig}'} &= \sqrt{\frac{\rho}{N}} \sum_{k, R_k \in \mathscr{L}^i} \widehat{\mathbf{h}}^i_{ik}\widehat{\mathbf{h}}^{i*}_{ik}\widehat{\mathbf{g}}^{i*}_{ik}\mathbf{g}^i_{ik} + \sqrt{\frac{\rho}{\rho'_{R,f}}} \mathbf{z}_f, && \text{(18a)} \\
h^{i,\text{int}'} &= \sqrt{\frac{\rho}{N}} \sum_{k, R_k \in \mathscr{L}^i} \widetilde{\mathbf{h}}^i_{ik}\widehat{\mathbf{h}}^{i*}_{ik}\widehat{\mathbf{g}}^{i*}_{ik}\mathbf{g}^i_{ik} - \sqrt{\frac{\rho}{\rho'_{R,f}}} \mathbf{z}_f \\
&\quad + \sqrt{\frac{\rho}{N}} \sum_{\substack{j=1 \\ j \neq i}}^{M} \sum_{k, R_k \in \mathscr{L}^j} \mathbf{h}^i_{jk}\widehat{\mathbf{h}}^{j*}_{jk}\widehat{\mathbf{g}}^{j*}_{jk}\mathbf{g}^i_{jk}. && \text{(18b)}
\end{aligned}$$

$\mathbf{y}^{i,\text{int}}$ and $\mathbf{v}^i$ are the same as in (16). From (16) and (18), note that the equality $h^{i,\text{sig}} = h^{i,\text{sig}'} + h^{i,\text{int}'}$ holds.

Now, we can apply the worst case noise theorem by showing that the data signal and the overall noise $\mathbf{V}^i$ are uncorrelated given CSI at the destination:

$$\text{E}\left\{\mathbf{s}^{i*}\mathbf{V}^i | h^{i,\text{sig}'}\right\} = 0.$$

$$\text{E}\left\{\mathbf{s}^{i*}\mathbf{V}^i | h^{i,\text{sig}'}\right\} = \text{E}\left\{\mathbf{s}^{i*}\left(\mathbf{s}^i h^{i,\text{int}'} + \mathbf{Z}^i\right) | h^{i,\text{sig}'}\right\}$$

$$= \text{E}\left\{\mathbf{s}^{i*}\mathbf{s}^i h^{i,\text{int}'} | h^{i,\text{sig}'}\right\}$$

$$+\mathrm{E}\left\{\mathbf{s}^{i*}\left[\sum_{\substack{j=1\\j\neq i}}^{M}\left(\mathbf{s}^{j}h_{j}^{i,\mathrm{int}}+\sum_{\substack{m=1\\m\neq i\\m\neq j}}^{N}\mathbf{s}^{m}h_{j,m}^{i,\mathrm{int}}\right)\right]\middle|h^{i,\mathrm{sig}'}\right\}$$

$$+\mathrm{E}\left\{\mathbf{s}^{i*}\mathbf{v}^{i}|h^{i,\mathrm{sig}'}\right\}.$$

The last term is zero since the data signal has zero mean, and $\mathbf{s}^i$ and $\mathbf{v}^i$ are independent. The second term is also zero since $\mathbf{s}^i$, $\mathbf{s}^j$, and $\mathbf{s}^m$ are independent, and have zero means. We rewrite the first term as follows since $\mathbf{s}^i$ and $h^{i,\mathrm{int}}$ are independent:

$$\mathrm{E}\left\{\mathbf{s}^{i*}\mathbf{s}^{i}h^{i,\mathrm{int}'}|h^{i,\mathrm{sig}'}\right\}$$
$$=\mathrm{E}\left\{\mathbf{s}^{i*}\mathbf{s}^{i}|h^{i,\mathrm{sig}'}\right\}\mathrm{E}\left\{h^{i,\mathrm{int}'}|h^{i,\mathrm{sig}'}\right\}.$$

The second term is zero since $\mathrm{E}\left\{\left(h^{i,\mathrm{sig}}-h^{i,\mathrm{sig}'}\right)|h^{i,\mathrm{sig}'}\right\}=0$. Therefore, we can replace the overall noise with a Gaussian noise. From (18), we evaluate the power ratio of the received desired signal and the overall noise in order to obtain the eSNR:

$$\rho_{\mathrm{eff,S2}}=\frac{\mathrm{var}\left\{\mathbf{y}^{i,\mathrm{sig}'}\right\}}{\mathrm{var}\{\mathbf{y}^{i,\mathrm{int}'}\}+\mathrm{var}\{\mathbf{y}^{i,\mathrm{int}}\}+\mathrm{var}\{\mathbf{v}^{i}\}},$$

which finally leads to (14). Note that we use the following property explained in Lemma 1 for evaluating (14):

$$\mathrm{var}\left\{\widehat{\mathbf{h}}_{ik}^{i}\widehat{\mathbf{h}}_{ik}^{i*}\widehat{\mathbf{g}}_{ik}^{i*}\mathbf{g}_{ik}^{i}\right\}=\mathrm{var}\left\{\widehat{\mathbf{h}}_{ik}^{i}\widehat{\mathbf{h}}_{ik}^{i*}\right\}\mathrm{var}\left\{\widehat{\mathbf{g}}_{ik}^{i*}\mathbf{g}_{ik}^{i}\right\}.$$

■

B. Effective SNRs of schemes S1a and S1b

The perfect and imperfect effective CSI at the destination are *provided by a genie* in S1a and S1b, respectively. On the other hand, the noisy imperfect effective CSI at the destination is provided by our *feedforwarding method* in S2 as shown in the previous subsection. Therefore, S1a and S1b are the *challenging* schemes of S2. In S1a, each relay has a perfect backward and forward CSI. In S1b and S2, each relay has an imperfect backward and imperfect forward CSI. We compare the eSNRs of these schemes numerically in the next subsection.

Since each relay has perfect CSI in S1a, each relay applies MF with the perfect backward and forward CSI. Therefore, for S1a, the received signal at the destination, $\mathbf{y}_{\mathrm{p}}^{i}$, is the same as $\mathbf{y}^{i}$ (16) if the imperfect CSI are replaced with the perfect CSI in $\mathbf{y}^{i}$:

$$\mathbf{y}_{\mathrm{p}}^{i}\triangleq\mathbf{y}^{i}(\widehat{h}=h,\widehat{g}=g),$$

Table III

SUMMARY FOR THE RELAY NETWORKS SCHEMES S1A, S1B, AND S2.

[1]CSI AT THE k^{th} RELAY IN THE i^{th} SUBGROUP.

[2]EFFECTIVE CSI AT THE i^{th} DESTINATION ANTENNA.

[3]RECEIVED SIGNAL AT THE i^{th} DESTINATION ANTENNA.

$i = 1, 2, \cdots, M$, AND $k, R_k \in \mathscr{L}^i$ $(k = 1, 2, \cdots, K)$.

	Relay CSI[1]	Dest. CSI[2]	Rec. sig.[3]	eSNR
S1a	$\mathbf{h}_{ik}^{i*}, \mathbf{g}_{ik}^{i*}$	$h_{\text{p}}^{i,\text{sig}}$	$\mathbf{y}_{\text{p}}^{i}$	$\frac{\text{var}\left\{\mathbf{y}_{\text{p}}^{i,\text{sig}}\right\}}{\text{var}\left\{\mathbf{y}_{\text{p}}^{i,\text{int}}\right\}+\text{var}\left\{\mathbf{v}_{\text{p}}^{i}\right\}}$
S1b	$\widehat{\mathbf{h}}_{ik}^{i*}, \widehat{\mathbf{g}}_{ik}^{i*}$	$h^{i,\text{sig}}$(16a)	$\mathbf{y}^i$(16)	$\frac{\text{var}\left\{\mathbf{y}^{i,\text{sig}}\right\}}{\text{var}\{\mathbf{y}^{i,\text{int}}\}+\text{var}\{\mathbf{v}^{i}\}}$
S2	$\widehat{\mathbf{h}}_{ik}^{i*}, \widehat{\mathbf{g}}_{ik}^{i*}$	$h^{i,\text{sig}'}$(18a)	$\mathbf{y}^i$(16)	$\frac{\text{var}\left\{\mathbf{y}^{i,\text{sig}'}\right\}}{\text{var}\{\mathbf{y}^{i,\text{int}'}\}+\text{var}\{\mathbf{y}^{i,\text{int}}\}+\text{var}\{\mathbf{v}^{i}\}}$

where p denotes the perfect CSI. Similarly, the perfect effective CSI for S1a, $h_{\text{p}}^{i,\text{sig}}$, is the same as $h^{i,\text{sig}}$ (16a) if the imperfect CSI are replaced with the perfect CSI in $h^{i,\text{sig}}$:

$$h_{\text{p}}^{i,\text{sig}} \triangleq h^{i,\text{sig}}(\widehat{h} = h, \widehat{g} = g).$$

Since $h_{\text{p}}^{i,\text{sig}}$ is available at the destination by a genie, the eSNR of S1a is as follows:

$$\rho_{\text{eff,S1a}} = \frac{\text{var}\left\{\mathbf{y}_{\text{p}}^{i,\text{sig}}\right\}}{\text{var}\left\{\mathbf{y}_{\text{p}}^{i,\text{int}}\right\} + \text{var}\left\{\mathbf{v}_{\text{p}}^{i}\right\}}.$$

Since each relay has imperfect CSI in S1b, each relay applies MF with the imperfect backward and forward CSI as in S2. Therefore, the received signal at the destination for S1b is exactly the same as S2, $\mathbf{y}^i$ (16). For S1b, the imperfect effective CSI, $h^{i,\text{sig}}$(16a), is also available at the destination by a genie. Therefore, the eSNR of S1b is as follows:

$$\rho_{\text{eff,S1b}} = \frac{\text{var}\left\{\mathbf{y}^{i,\text{sig}}\right\}}{\text{var}\{\mathbf{y}^{i,\text{int}}\} + \text{var}\{\mathbf{v}^{i}\}}.$$

As a result,

$$\rho_{\text{eff,S1a}} > \rho_{\text{eff,S1b}} > \rho_{\text{eff,S2}}.$$

Clearly, the order of the effective CSI from more accurate to less accurate is $h_{\text{p}}^{i,\text{sig}}$, $h^{i,\text{sig}}$(16a), and $h^{i,\text{sig}'}$(18a), which are available at the destination for S1a, S1b, and S2, respectively. We omit the explicit derivations and the expressions of $\rho_{\text{eff,S1a}}$ and $\rho_{\text{eff,S1b}}$ for the sake of brevity. The summary of the schemes is presented in Table III.

C. Numerical comparison

We compare the schemes S1a, S1b, and S2 when $M = 2$ and $K = 7$ or $K = 20$ while N varies from 1 to 6. We choose $M = 2$ since it is the least number expected for the MIMO technology. We choose $K = 7$ and $K = 20$ to clearly demonstrate the effect of the increase in the number of relays. For convenience, we use the simplified versions of the eSNRs by assuming that $\sigma^2_{\widehat{h}} = \sigma^2_{\widehat{h}^y_x}$ and $\sigma^2_{\widehat{g}} = \sigma^2_{\widehat{g}^y_x}$ $(x \in (i, j)$ and $y \in (i, j, m))$. Finally, we assume that the channel variances are equal to unity. Note that assuming equal channel estimation variances *within* the backward and forward channels does not limit or bias our observation since we do not use a feature regarding this.

The assumption of unequal channel estimation variances *among* the backward and forward channels is important since we use the MF method, which performs accordingly. A relay network with the MF method has an asymmetrical structure in nature. For the forward channels, the MF is a pre-filtering before the noise corruption. On the other hand, for the backward channels, the MF is just a rough estimation of the data signal due to the post-filtering after the noise corruption. Therefore, the effects of the backward and forward channel estimations on the eSNR are asymmetrical. In order to show this asymmetric effect on the eSNR, we evaluate the eSNRs by changing the backward (forward) channel estimation variance discretely from 0.1 to 0.9 in 0.4 increments while holding the forward (backward) channel estimation variance at 0.9. Since the MF is used at the relays, the eSNR with a more accurate backward channel estimate is higher than the eSNR with a more accurate forward channel estimate as seen in Fig. 4 and Fig. 5. In other words, for the MF method, having more accurate backward channel estimates is more valuable than having more accurate forward channel estimates.

Realistic K assumption: Our results show the importance of considering the parameter K realistically. Although we considerably increase K from 7 to 20, the eSNRs of all schemes are still far from converging. If we unrealistically assume $K \to \infty$, the eSNRs of all schemes converge to K as also seen in [6].

Feedforwarding method: The scheme with the feedforwarding method S2 has unsatisfactory performance due to its simplicity when we *directly* compare it with S1a and S1b. However, our scheme S2 can still achieve a higher eSNR than the scheme without the channel estimation error (S1a) at the cost of one additional antenna at each relay as seen in Fig. 4 and Fig. 5. This significant achievement of the feedforwarding method is due to its ability of providing a perfect forward CSI at the destination (e.g., perfect forward CSI g_k as seen in (11) for the single antenna case). Fixed relays can support a level of intelligence and cost budget between the levels of the mobile nodes and the base stations can support [10]. Therefore, for a network with fixed relays, it is reasonable to invest more at the relays.

(a) More accurate backward channel estimation.

(b) More accurate forward channel estimation.

Figure 4. eSNRs of the schemes S1a, S1b, and S2 when $K = 7$.

V. CONCLUSION

In this paper, we introduced a new training protocol for wireless relay networks, which provides a higher eSNR (thus a higher capacity) and a faster transfer of the CSI than the existing training protocols. In addition, unlike the other protocols, the CSI transfer time of our protocol is independent of the number of relays K, which is important for large K values. We also showed the importance of the effective CSI accuracy on the eSNR by presenting the numerical eSNR results of the networks with perfect effective CSI, imperfect effective CSI, and noisy imperfect effective CSI available at the destination.

(a) More accurate backward channel estimation.

(b) More accurate forward channel estimation.

Figure 5. eSNRs of the schemes S1a, S1b, and S2 when $K = 20$.

VI. ACKNOWLEDGMENT

The authors would like to thank S. Murugesan and P. Schniter at The Ohio State University for helpful discussions during the early stages of this research. In addition, they would like to thank the anonymous reviewers for their helpful comments and suggestions that significantly improved the quality and the presentation of the paper.

References

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Process. Mag.*, pp. 12–25, Nov. 2004.

[2] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - part i: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[4] ——, "User cooperation diversity - part ii: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[5] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.

[6] H. Bolcskei, R. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[7] B. Wang, J. Zhang, and L. Zheng, "Achievable rates and scaling laws of power-constrained wireless sensory relay networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4084–4104, Sep. 2006.

[8] V. I. Morgenshtern and H. Bolcskei, "Crystallization in large wireless networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3319–3349, Oct. 2007.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[10] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16J relay-based wireless access networks: An overview," *IEEE Wireless Commun. Mag.*, vol. 15, no. 5, pp. 56–63, Oct. 2008.

[11] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[12] ——, "Towards an information theory of large networks: an achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[13] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[14] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[15] I. S. Reed, "On a moment theorem for complex gaussian processes," *IEEE Trans. Inf. Theory*, vol. 8, no. 3, pp. 194–195, Apr. 1962.

[16] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May. 2008.

[17] R. Ware and F. Lad, "Approximating the distribution for sums of products of normal variables," University of Canterbury, http://www.math.canterbury.ac.nz/php/research/abstracts/abstract9.php, Tech. Rep., 2003.